

Client Report - What's in a Name?

Unit 1 Task 2

AUTHOR

Dawson Packer

```
# Bring in pandas to read the spreadsheet-like data
import pandas as pd
# Bring in Markdown so I can print a short written answer
from IPython.display import Markdown
# Bring in the plotting pieces I need from lets_plot
from lets_plot import ggplot, aes, geom_line, geom_vline, labs, theme_minimal, scale_x_continuous
# Turn on lets_plot visuals inside this report
LetsPlot.setup_html(isolated_frame=True)
# Store my own name so I can use it in Question 1
preferred_name = "Dawson"
# Store my birth year so I can mark it on the chart
birth_year = 2003
```

Data Preparation

```
# Read the SSA name counts straight from the class GitHub file
names_year = pd.read_csv("https://github.com/byuidatascience/data4names/raw/master/data-raw/names_
```

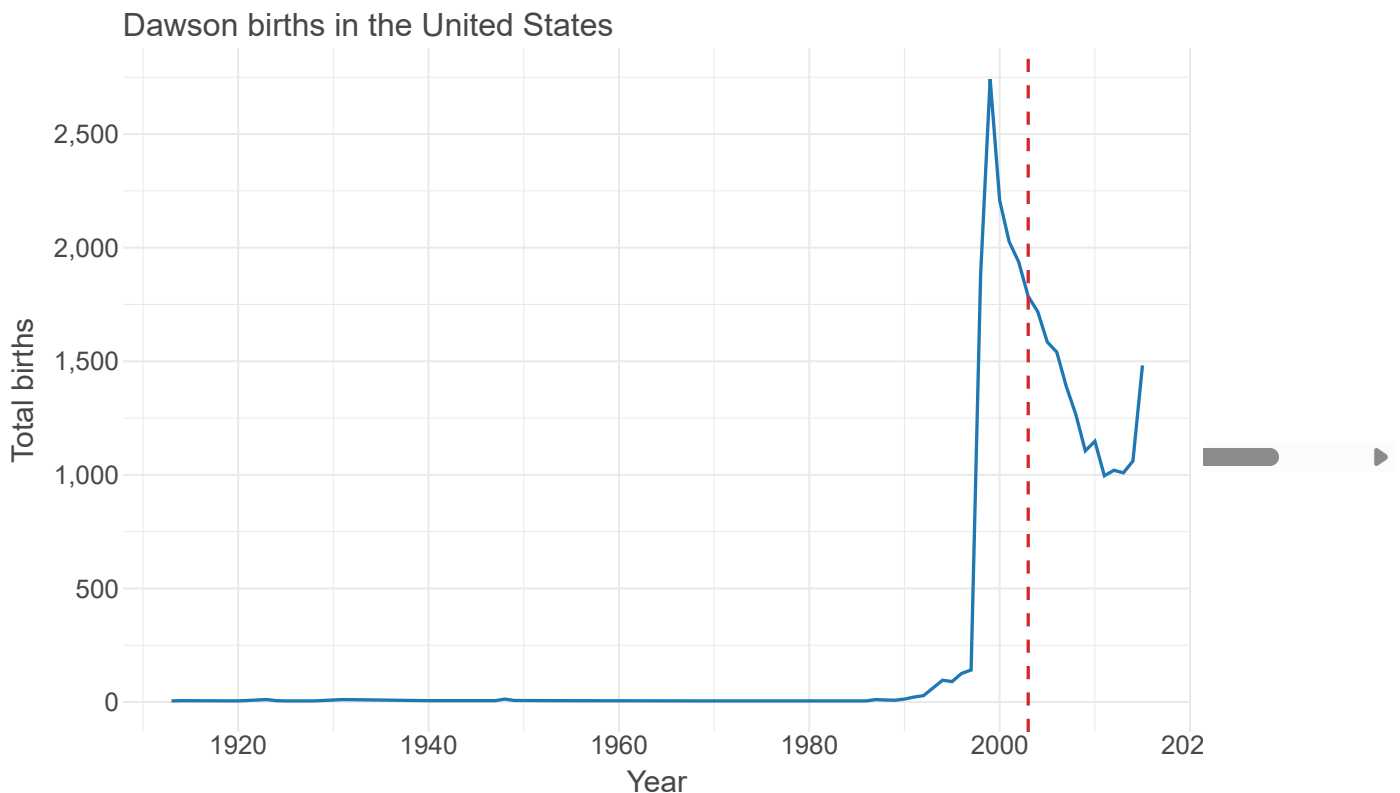
QUESTION 1

How does your name at your birth year compare to its use historically?

```
# Filter the table to just my name and sort the years so the line looks right
my_history = names_year.loc[names_year["name"] == preferred_name, ["year", "Total"]].sort_values('year')
# Build a simple line chart with a dashed line at my birth year
plot_q1 = (
    ggplot(my_history, aes("year", "Total"))
    + geom_line(color="#1f77b4")
    + geom_vline(xintercept=birth_year, linetype="dashed", color="#d62728")
    + scale_x_continuous(format='d')
    + labs(title=f"{preferred_name} births in the United States", x="Year", y="Total births")
    + theme_minimal()
)
# Pull the total for my birth year so I can mention it
birth_total = int(my_history.loc[my_history["year"] == birth_year, "Total"].iloc[0])
# Pull the year where the name hit its national peak
peak_row = my_history.loc[my_history["Total"].idxmax()]
# Write a one-sentence takeaway comparing my year with the peak
```

```
note_q1 = (
    f"In {birth_year}, {birth_total:,} babies were named {preferred_name}. "
    f"The peak year was {int(peak_row['year'])} with {int(peak_row['Total']):,} births."
)
```

```
# Show the Question 1 chart
plot_q1
```



```
# Share the quick explanation for Question 1
Markdown(note_q1)
```

In 2003, 1,785 babies were named Dawson. The peak year was 1999 with 2,742 births.

QUESTION 2

If you talked to someone named Brittany on the phone, what is your guess of his or her age? What ages would you not guess?

```
# Filter the table to Brittany and sort the years so the line reads left to right
brittany_history = names_year.loc[names_year["name"] == "Brittany", ["year", "Total"]].sort_values
# Draw Brittany's national totals with a dashed line at the peak year
peak_year = int(brittany_history.loc[brittany_history["Total"].idxmax(), "year"])
plot_q2 = (
    ggplot(brittany_history, aes("year", "Total"))
```

```

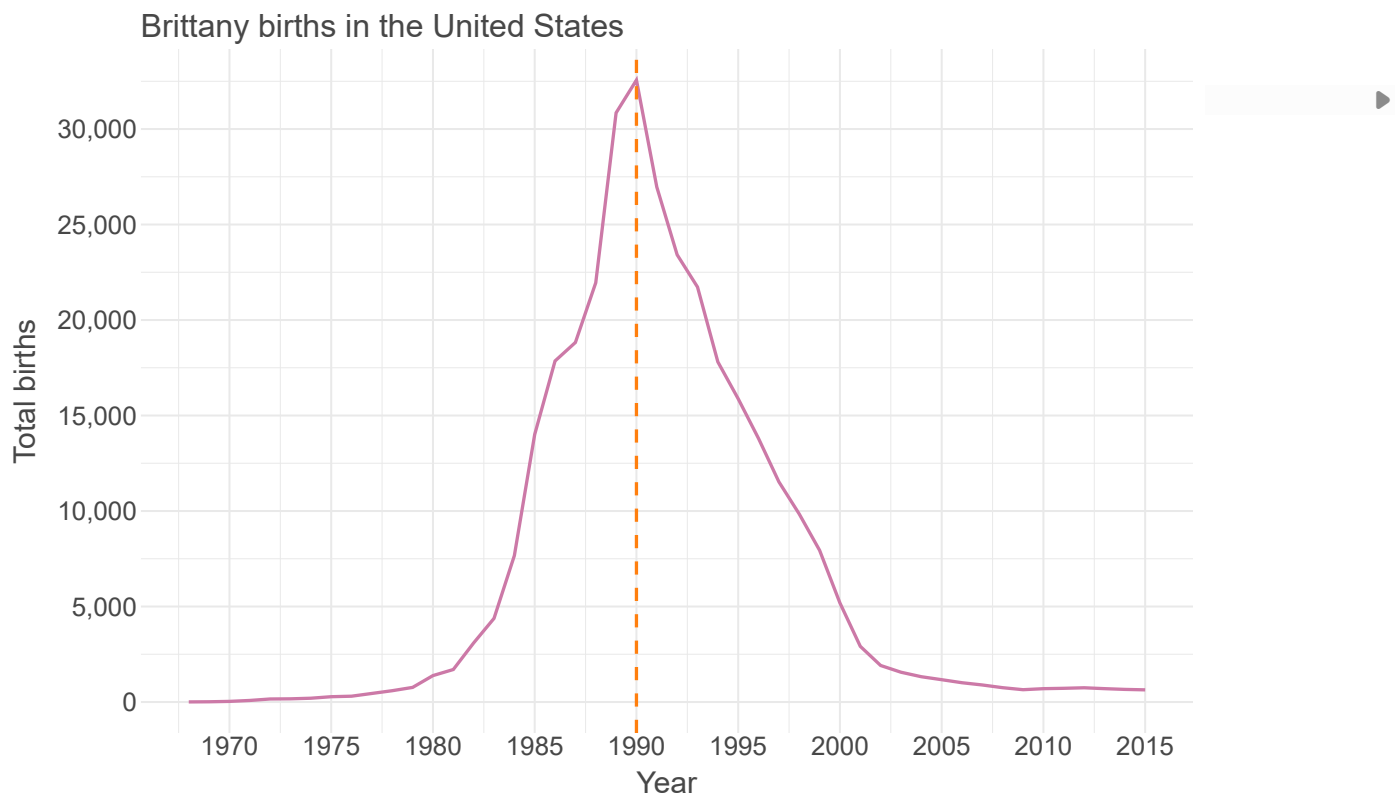
+ geom_line(color="#cc79a7")
+ geom_vline(xintercept=peak_year, linetype="dashed", color="#ff7f0e")
+ scale_x_continuous(format='d')
+ labs(title="Brittany births in the United States", x="Year", y="Total births")
+ theme_minimal()
)
# Take a weighted average birth year so I can turn it into an age guess
guess_year = int(round((brittany_history["year"] * brittany_history["Total"]).sum() / brittany_his
# Convert to an age range using the current year (2025)
age_low, age_high = 2025 - peak_year, 2025 - guess_year
# Mention the years with the lowest usage so we can say which ages we would avoid
early_year = int(brittany_history.loc[brittany_history["Total"] == brittany_history["Total"].min()
late_year = int(brittany_history.loc[brittany_history["Total"] == brittany_history["Total"].tail(
# Write a plain sentence explaining the guess and the ages to avoid
note_q2 = (
    f"Brittany peaked in {peak_year} and averages around {guess_year}, so I would guess she is {age
    f"I would avoid guessing ages tied to the very early years like {early_year} or the late-year
)

```

```

# Show the Brittany chart
plot_q2

```



```

# Share the Brittany age estimate
Markdown(note_q2)

```

Brittany peaked in 1990 and averages around 1991, so I would guess she is 34 or 35. I would avoid guessing ages tied to the very early years like 1968 or the late-year dip in 2015 because almost no babies were

named Brittany then.